

(12) UK Patent Application (19) GB (11) 2 312 582 (13) A

(43) Date of A Publication 29.10.1997

(21) Application No 9601038.8

(22) Date of Filing 19.01.1996

(71) Applicant(s)

Orod Hi-Tech Systems Limited

(Incorporated in Israel)

PO BOX 2177, 14 Herta's St, Savir CTR,  
Kfar Saba 44425, Israel

(72) Inventor(s)

Avi Sharir  
Michael Tamir  
Razhok Wilf  
Shmuel Patog

(74) Agent and/or Address for Service

Eric Potter Clarkson  
St Mary's Court, St Mary's Gate, NOTTINGHAM,  
NG1 1LE, United Kingdom

(51) INT CL<sup>6</sup>

H04N 5/265

(52) UK CL (Edition O )

H4F FD2B FD30K FGS

(56) Documents Cited

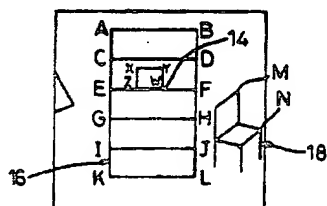
WO 91/15921 A1

(58) Field of Search

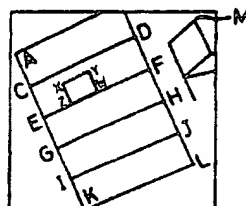
UK CL (Edition O ) H4F FGG FGH FGJ FGS FGT  
INT CL<sup>6</sup> G05T 11/00 , H04N 5/265 5/272 9/74 9/78  
Online: WPI, INSPEC

(54) Insertion of virtual objects into a video sequence

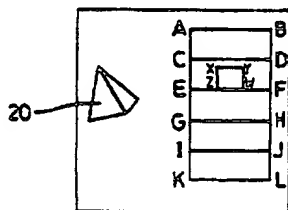
(57) A computer generated character is inserted into a sequence of video frames, the sequence having a set of feature points A-L in the first 1A, last 1C and intermediate 1B frames, by manually inserting the computer generated character 14, 1D into the first and last frames of the sequence, the position of the character in each intermediate frame of the sequence being determined by automatic calculation using the feature points A-L and reference points X-Z on the computer generated character.



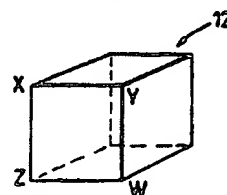
**Fig. 1A**  
FRAME I



**Fig. 1B**  
FRAME K

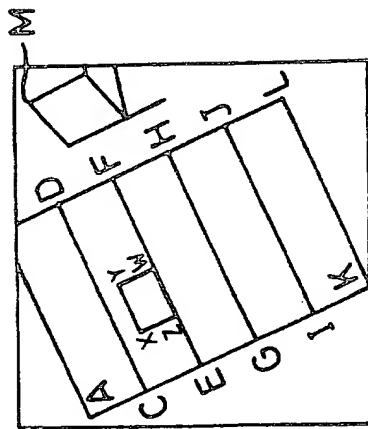


**Fig. 1C**  
FRAME N

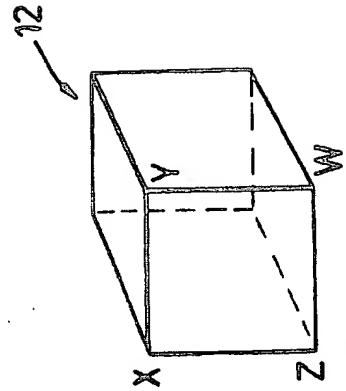


**Fig. 1D**  
VIRTUAL OBJECT

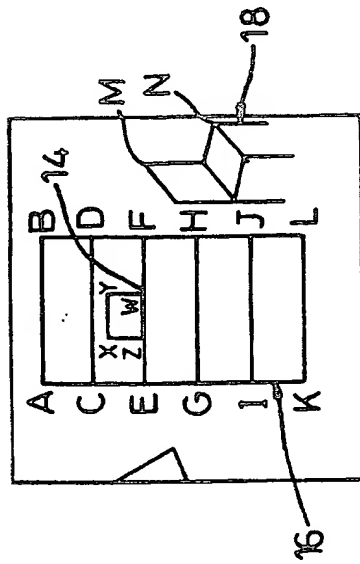
GB 2 312 582 A



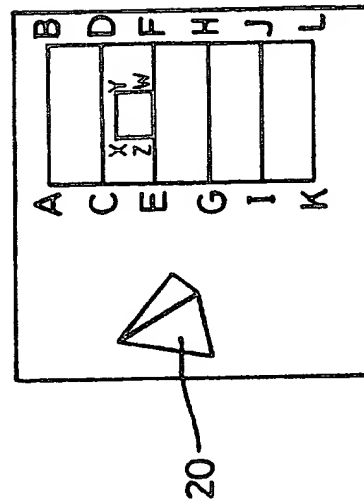
**Fig. 1B**  
FRAME K



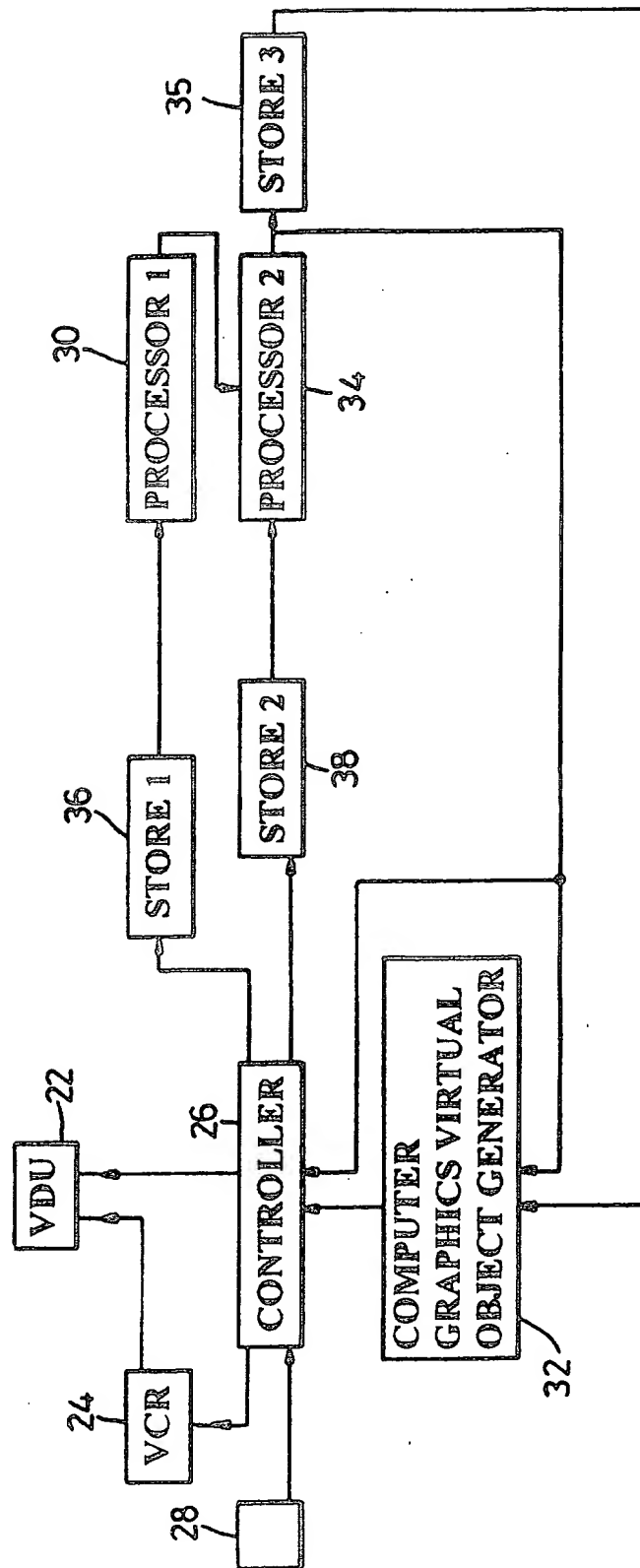
**Fig. 1D**  
VIRTUAL OBJECT

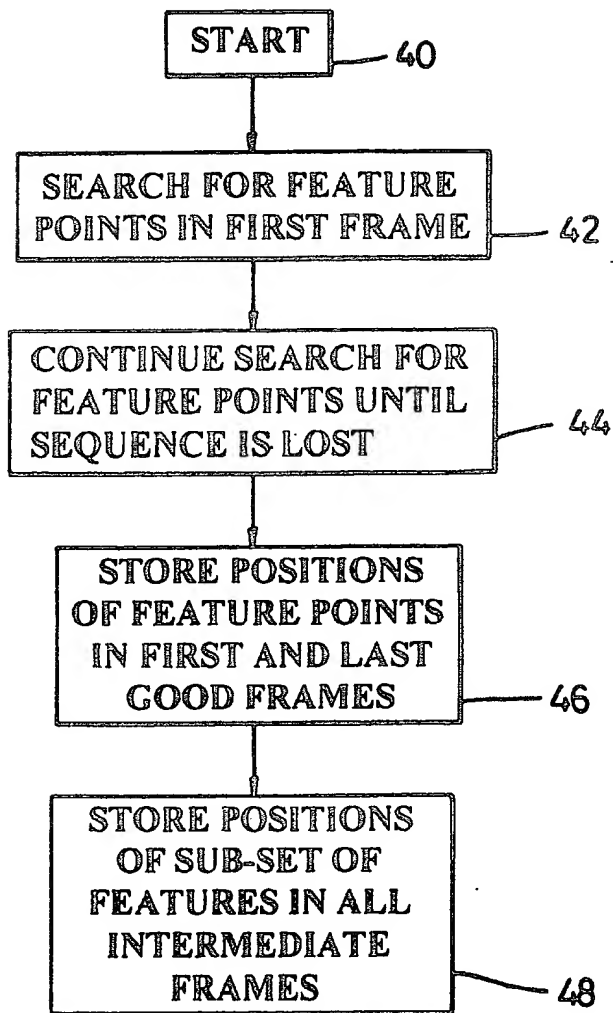


**Fig. 1A**  
FRAME I

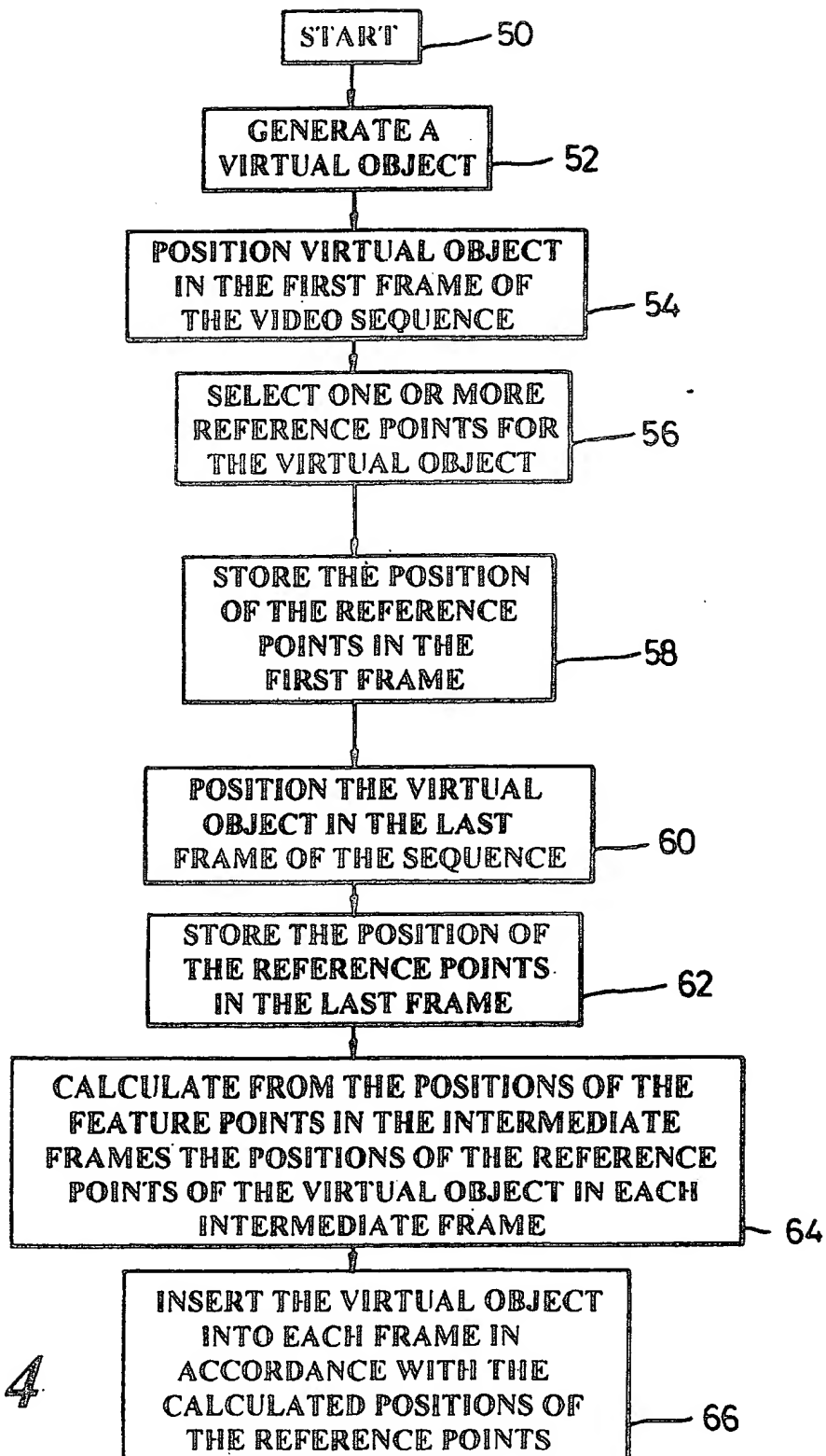


**Fig. 1C**  
FRAME N

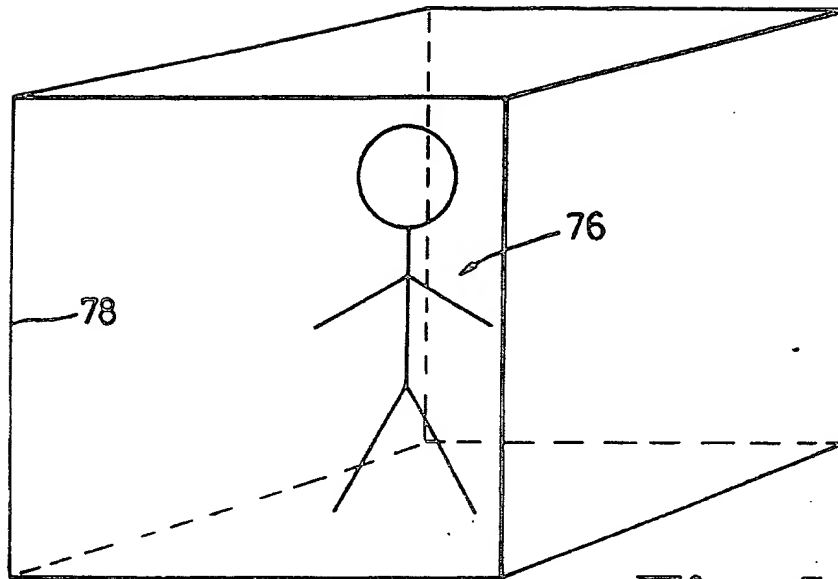
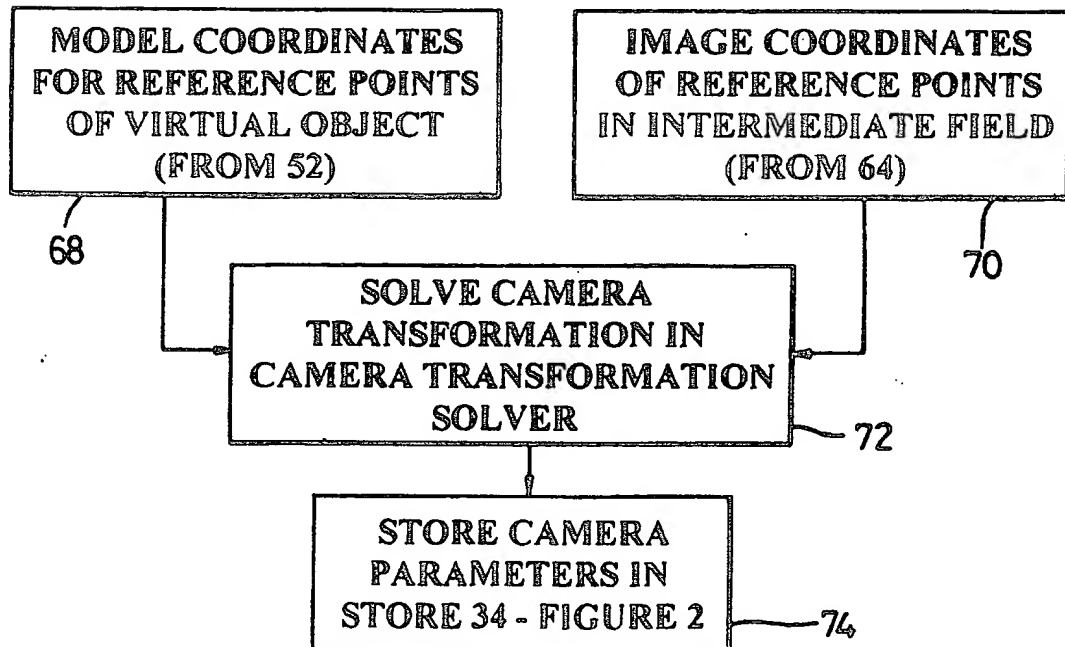
*Fig. 2*

*Fig. 3*

4/5



*Fig. 4*

*Fig. 5**Fig. 6*

METHOD AND APPARATUS FOR INSERTION OF VIRTUAL  
OBJECTS INTO A VIDEO SEQUENCE

5 The present invention relates to insertion of virtual objects into video sequences and in particular to sequences which have already been previously generated.

10 Computer generated (CG) images and characters are widely used in feature films and commercials. They provide for special effects possible only with CG content as well as for the special look of a cartoon character. While in many instances the complete picture is computer generated, in other instances, CG characters are to be inserted in a live image sequence taken by a physical camera.

15 Prior art describes how CG objects are inserted in a background photograph, for the purpose of architectural simulation [E Nakamae et al., A montage method: the overlaying of the computer generated images onto a background photograph, ACM Trans. on Graphics, Vol. 20, No. 4, 1986 (207-241)]. That method solves the viewpoint from a set of image points, matched with their geographical map locations. In other practical  
20 situations, no measured three-dimensional data can be associated with image points. Therefore, insertion is done manually, using a modeller to transform the CG object, until it is registered with the image.

25 Consider the automatic insertion of three dimensional virtual objects in image sequences. While manual techniques are suitable for a single picture, they pose practical problems when processing a sequence of images:

- A typical shot of a few seconds involves hundreds of images, making the manual work tedious and error-prone.
- Independently inserting the CG objects at each image might  
30 introduce spatial jitter over time, although the insertion may look

perfect at each frame.

In a real motion picture, the apparent motion of the objects and the characters is a combination of the objects ego-motion in a 3D world, and the motion of the camera.

5        For CG characters, the ego-motion is determined by the animator. Then, camera motion has to be applied to the characters.

One possible solution is to use motion control systems in shooting the live footage. In such systems, the motion of the camera is computer-controlled and recorded. These records are then used in a straight  
10 forward manner to render the CG characters in synchronization with camera motion.

However, in many practical cases, the usage of motion control systems is inconvenient.

If a known 3D object is present in the sequence, it may be used to  
15 solve camera motion, by matching image features to the object's model. If this is not the case, we may try to solve the structure and the motion concurrently [J Weng et al., Error Analysis of Motion Parameter Estimation from Image Sequences, First Intl. Conf. on Computer Vision 1987, pp. 703-707]. These non-linear methods are inaccurate, slowly  
20 converging and computationally unstable.

One may note that for the application at hand, we have no use for the explicitly camera model other than projecting the virtual object, at each view of the sequence, using the corresponding camera model. Thus, in the present invention we suggest merging the 3D estimation and  
25 projection stages into one process which predicts the image-space motion of the virtual object from image-space motion of tracked features.

The present application provides a method and apparatus for insertion of CG characters into a existing video sequence, independent of motion control records or a known pattern.

30



According to the present invention there is provided a method of insertion of virtual objects into a video sequence consisting of a plurality of video frames comprising the steps of :

- 5           i.     detecting in a one frame (Frame A) of the video sequence a set of feature points;
- ii.    detecting in another frame (Frame B) of the video sequence the set of feature points;
- iii.   detecting in each frame other than frame A or frame b at least a sub-set of the feature points;
- 10          iv.   positioning a virtual object in a defined position in frame A;
- v.     positioning the virtual object in the defined position in frame B;
- vi.    selecting one or more reference points for the virtual object;
- vii.   computing the position of the reference points in each frame
- 15          viii. inserting the virtual object in each frame in the position of the sequence; and determined by the computation.

According to a further aspect of the present invention there is provided apparatus for insertion of virtual objects into a video sequence

20   consisting of a plurality of video frames said apparatus including :

- i.     means for detecting in one frame (Frame A) a set of feature points;
- ii.    means for detecting in another frame (Frame B) the set of feature points;
- 25          iii.   means for detecting in each frame other than frame A or frame B at least a sub-set of the feature points;
- iv.    means for positioning a virtual object in a defined position in frame A;
- v.     means for positioning the virtual object in the defined
- 30          position in frame B;

- vi. means for selecting one or more reference points for the virtual object;
- vii. means for computing the position of the reference points in each frame of the sequence; and
- 5 viii. means for inserting the virtual object in each frame in the position determined by the computation.

In a preferred embodiment of the present invention, the CG character is constrained relative to a cube or other regularly shaped box, the cube representing the virtual object. The CG character is thereby able  
10 to be animated.

The present invention will now be described, by way of example with reference to the accompanying drawings, in which :

15 Figure 1 shows an exemplary video sequence illustrating in Figure 1A a first frame of the video sequence; in Figure 1B an intermediate frame (K) of the video sequence; in Figure 1C a last frame of the video sequence and in Figure 1D a virtual object to be inserted into the video sequence of Figures 1A to 1C;

20 Figure 2 shows apparatus according to the present invention;

Figure 3 shows a flow diagram illustrating the selection and storage of feature points;

Figure 4 shows a flow diagram illustrating the positioning of the virtual object in the first, last and intermediate frames;

25 Figure 5 shows a cube (as defined) enclosing a three dimensional moving virtual character; and

Figure 6 shows a flow diagram illustrating the solution of camera transformation corresponding to a frame.

The present invention is related to the investigation properties of  
30 feature points in three perspective views, As an example, consider the

concept of the fundamental matrix (FM). [R Deriche et al., Robust recovery of the epipolar geometry for an uncalibrated stereo rig, Lecture Notes in Computer Science, Vol. 800, Computer Vision - ECCV 94, Springer-Verlag Berlin Heidelberg 1994, pp. 567-576]. Given 2  
5 corresponding points in two views,  $q$  and  $q'$  (in homogeneous coordinates) we can write :

$$q'Fq = 0$$

where the  $3 \times 3$  matrix  $F$ , which describes this correspondence is known as the fundamental matrix. Given 8 or more matched point pairs, we can in  
10 general determine a unique solution for  $F$ , defined up to a scale factor.

Now, consider 3 images with two corresponding pixels  $m_1$  and  $m_2$  in images 1 and 2. Where should the corresponding pixel  $m_3$  be in picture 3? Let  $F_{13}$  be the fundamental matrix of images 1,3 and let  $F_{23}$  be the fundamental matrix of images 2,3. Then  $m_3$  is given by the intersection  
15 of the epipolar lines:

$$F_{13}m_1, F_{23}m_2$$

[O. Faugeras and L. Robert, What can two images tell us about a third one? Lecture Notes in Computer Science, Vol. 800, Computer Vision - ECCV 94, Springer-Verlag Berlin Heidelberg 1994, pp. 485-492]

20 The fundamental matrix is used later in the description of the embodiment of the invention. However, the invention is not limited to this specific implementation. Other formulations could be used, for example the concept of tri-linearity (TT) [A. Shashua and M. Werman, Trilinearity of three perspective views and its associated tensor, IEEE 5th  
25 Intl. Conf. on Computer Vision, 1995, pp. 920-925]

Specific embodiments of the invention are now described with reference to the accompanying figures.

With reference now to Figure 1, Figure 1A shows a first video frame which is assumed to be the first frame of a sequence, selected as  
30 now described.

The sequence can be selected manually or automatically. For each sequence either the operator or an automatic feature selection system searches for a number of feature points in both a first frame (Frame 1) Figure 1A and a last frame (Frame N) Figure 1C. In any intermediate  
5 frame such as in Figure 1B (Frame K) a sub-set of the points must be visible. In a preferred embodiment there should be at least 8 (eight) feature points in all intermediate frames since this can be used both for the FM methods which requires at least 8 points along three frames, or the TT method which requires at least 7 points along three frames.

10 In frame 1 (Figure 1A) feature points A-L (12 points) are recognised. In Figure 1B where the camera has tilted and possibly zoomed point B is missing. In Figure 1C all 12 points are again visible. It is noted that in Figure 1A a chair M,N is visible, this being also visible in Figure 1B but not Figure 1C. This chair M.N is not used for  
15 calculation.

An object (Figure 1D) is computer generated and in this example comprises a cube 12 (XYZW). The cube 12 is to be positioned on a shelf 14 of a bookcase 16.

In the first scene of the video sequence a chair 18 is shown, but  
20 although the chair 18 is present in the intermediate frame (K) Figure 1B it is not present in the last frame in the sequence. Thus it is not used to define points. Similarly cone 20 is present in the last frame but not in the first or Kth frame. Thus this cone 20 is not used but only the bookcase 16 is used.

25 In Figures 1A and 1C all corners of the shelves are visible (A-L). In Figure 1B only 11 out of 12 corners are visible since corner B is missing. However, in all video frames A-L at least a minimum number of feature points are visible. In a preferred embodiment this minimum number is eight and these must be visible in all frames.

30 With reference now to Figure 2, the VDU 22 receives a video

sequence from VCR 24. The video controller 26 can control VCR 24 to evaluate a sequence of video shots as in Figures 1A to 1C to evaluate a sequence having the desired number of feature points. Such a sequence could be very long for a fairly static camera or short for a fast panning  
5 camera.

The feature points may be selected manually by, for example, mouse 28 or automatically. Preferably, as stated above, at least eight feature points are selected to appear in all frames of a sequence. When the controller 26 in conjunction with processor 30 detects that there are  
10 less than eight points the video sequence is terminated. If further insertion of an object is required then a continuing further video sequence is generated using the same principles.

Assuming therefore that the sequence of video frames I to N has been selected, a computer generated (CG) object 12 is created by  
15 generator 32. The CG object 12 is then positioned as desired in the first and last frames of the sequence. The orientation of the object in the first and last frames is accomplished manually such that the object appears to be naturally correct in both frames. The CG object 12 is then automatically positioned in all intermediate frames by the processors 30  
20 and 34 as follows with reference to Figures 3 and 4.

From a start 40 processor 1 searches for feature points in a first frame 42 and continues searching for these features until the sequence is lost 44. The feature positions are then stored in store 36 - step 46. The positions of these features in all intermediate frames are then stored in  
25 store 36 - step 48.

The CG object 12 is then generated 50, 52 - Figure 4 and positioned on the shelf 14 in a first frame of the video sequence - step 54.

One or more reference points are selected for the CG object - step 56. These could be the four non co-planar corners of the cube 12 or  
30 could be other suitable points on an irregularly shaped object.

The positions of the reference points in the first frame are stored in store 38 - step 58.

The CG object is then positioned in the last frame of the sequence - step 60 and the position of the reference points is stored for this position of the CG object in store 38 - step 62.

Using both processor 30 and processor 34 the positions of the reference points for the object 12 are calculated for each intermediate frame  $i$  by calculating the FM or the TT using the triplets of reference points in the first frame, last frame and frame  $i$  - step 64. The location of the reference points for the object in Frame  $i$  is computed from the locations of the corresponding object points in the first and in the last frames, as well as the FM or the TT as described before.

From these positions the virtual CG object 12 is inserted into each frame in accordance with the calculated positions of the reference points - step 66. The insertion is carried out by controller 26 under the control of inputs from processor 34 and from graphics generator 32 also controlled by processor 32. Alternatively, compute the TT of the first, last, and intermediate frames using at least 7 corresponding feature points in the three frames.

The process described in Figures 3 and 4 comprises a virtual point prediction using the fundamental matrix or the TT.

In Figures 3 and 4 we:

1. Position the virtual object in the first (1) and last frame (2):
2. For each frame except the first and last frames:
  - 2.1 use at least 8 corresponding feature points to compute the fundamental matrix  $F_{1k}$  between the first and intermediate frame: use at least 8 corresponding feature points to compute the fundamental matrix  $F_{Nk}$  between the last and intermediate frame.
  - 2.2 For each reference point (to be predicted) such that its

location in the first frame (1 as determined by process 52 is  $m_1$  and its location in frame N is  $m_N$ ; compute the lines  $F_{1K} m_1, F_{NK} m_N$ . Intersect the lines to obtain  $m_K$ .

Alternatively, the location of the reference point  $m_1$  can be  
5 computed using the TT and its locations in the first and in the last frames.

If, as shown by way of a preferred example the CG object is a cube or other regular solid shape (hereinafter referred to as a cube) there is a possibility of providing an animated figure which is associated with the  
10 cube. The figure may be completely within the cube or could be larger than the cube but constrained in its movement in relation to the cube.

Since the cube is positioned relative to the video sequence the animated figure will also be positioned. Thus, if for example the cube  
15 was made a rectangular box which was the size of shelf 16, then a rabbit could be made to dance along the shelf.

It may be seen therefore that the example described in Figure 4, is a complete recipe for wire-frame virtual objects, since it allows to compute the position of all vertices at each intermediate from.

20

However, this solution is not complete for most practical cases, where surface rendering and object ego-motion are required. For these cases we must derive a three-dimensional virtual object description at each frame.

25 We now describe how we deal with surface rendering and ego-motion.

In step 54 when we position the virtual object the transformation applied to the model in 52 can be stored and the inverse of the transformation constitutes a camera transformation due to the duality  
30 between the camera and object motions.

Therefore when we generate the virtual object in 52 we would prefer to generate it relative to a rectangular bounding box (see Figure 5) and then the vertices of this bounding box can be used as reference points in 64.

5        Given the position of the reference points in the intermediate frames, the camera transformation corresponding to the frame can be solved as indicated in Figure 6 in which in step 68 the model coordinates for the reference points of the virtual object from step 52 of Figure 4 are used with the image coordinates of reference points in the intermediate  
10    field (step 70) to combine to solve camera transformation information (step 72) and this is then stored in store 35 (Figure 1) - step 74.

Solving camera transformation information from image coordinates of reference points is described in [C.K. Wu et al., Acquiring 3-D spatial data of a real object, Computer Vision, Graphics and Image Processing  
15    28, 126-133 (1980)].

Now, with reference to Figure 5, this transformation is applied to the actual object so that if we allow the virtual character 76 to move relative to the bounding box 78 in the object coordinate system then we take the animated model (character) at each intermediate frame and further  
20    transform it by the camera transformation computed as described above.

The animated model will therefore move naturally and the correct perspective etc will be provided by the camera transformation system as calculated above.

An alternative method to insert an object having ego motion is to  
25    generate it manually only in the coordinate systems of frame A and frame B. This can be manually adjusted by an animator for correct appearance in both images. The entire object can then be reprojected into all other frames by using its locations in Frames A and B, and the FM or TT methods.

30



## CLAIMS

1. A method of insertion of virtual objects into a video sequence consisting of a plurality of video frames comprising the steps of :
  - 5 i. detecting in a one frame (frame A) of the video sequence a set of feature points;
  - ii. detecting in another frame (frame B) of the video sequence the set of feature points;
  - iii. detecting in each frame other than frame A or frame B at  
10 least a sub-set of the feature points;
  - iv. positioning a virtual object in a defined position in frame A;
  - v. positioning the virtual object in the defined position in frame B;
  - vi. selecting one or more reference points for the virtual object;
  - 15 vii. computing the position of the reference points in each frame of the sequence; and
  - viii. inserting the virtual object in each frame in the position determined by the computation.
- 20 2. A method as claimed in claim 1 in which the computation of the position of the reference points (step vii) is carried out by calculation of the positions of the feature points in each intermediate frame and by geometric transformation of the position of the reference points in relation to the feature points.
- 25 3. A method as claimed in claim 1 or claim 2 in which the virtual object is represented by a box, the reference points being corners of the box.
- 30 4. A method as claimed in claim 3 in which a virtual character is

positioned within or in fixed relationship to the box.

5. A method as claimed in claim 4 in which the virtual character is animated.

5

6. A method as claimed in any one of claims 1 to 5 in which the set of feature points is selected automatically.

7. A method as claimed in anyone of claims 1 to 6 in which the  
10 computation of the position of the feature points is carried out by tracking of each feature point on a frame by frame basis.

8. Apparatus for insertion of virtual objects into a video sequence consisting of a plurality of video frames said apparatus including :

- 15 i. means for detecting in one frame (frame A) a set of feature points;
- ii. means for detecting in another frame (frame B) the set of feature points;
- 20 iii. means for detecting in each frame other than frame A or frame B at least a sub-set of the feature points;
- iv. means for positioning a virtual object in a defined position in frame A;
- v. means for positioning the virtual object in the defined position in frame B;
- 25 vi. means for selecting one or more reference points for the virtual object;
- vii. means for computing the position of the reference points in each frame of the sequence; and
- 30 viii. means for inserting the virtual object in each frame in the position determined by the computation.

9. Apparatus as in claim 8 including means for representing the virtual object.
- 5 10. Apparatus as claimed in claim 9 including means for positioning a virtual character within a rectangular box.
11. Apparatus as claimed in claim 10 including means for animating the virtual character.
- 10 12. Apparatus as claimed in claim 8 including means for automatically selecting the set of feature points.
- 15 13. Apparatus as claimed in claim 8 in which the means for computation of the position of the feature point comprises means for tracking of each point on a frame by frame basis.



The  
Patent  
Office

Application No: GB 9601098.8  
Claims searched: All

Examiner: Joe McCann  
Date of search: 17 April 1996

**Patents Act 1977**  
**Search Report under Section 17**

**Databases searched:**

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK CI (Ed.O): H4F(FGG,FGH,FGJ,FGS,FGT)

Int CI (Ed.6): H04N(5/262,5/272,9/74)

Other: Online: WPI, INSPEC

**Documents considered to be relevant:**

Category	Identity of document and relevant passage	Relevant to claims
X	WO 91/15921 (MULTI-MEDIA TECHNIQUES) - see abstract	1,8

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**